


Research Article

LDV-YOLO: A Lightweight Improved YOLOv11 for Photovoltaic Cell Defect Detection

Zhihui Li¹, Liqiang Wang^{1,2*} , Dong You¹, Lirong Chen¹

¹School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China

²Tianjin Engineering Research Center of Fieldbus Control Technology, Tianjin, 300202, China
E-mail: wangliqiang@tute.edu.cn

Received: 19 June 2025; **Revised:** 17 July 2025; **Accepted:** 23 July 2025

Abstract: Currently, the detection of defects in solar cells faces several challenges, including the difficulty in identifying small-sized defects, high sensitivity to complex backgrounds, and the large size of existing models. To overcome these limitations, this study proposes Laser Doppler Velocimetry-You Only Look Once (LDV-YOLO), a lightweight model for efficient defect detection. Initially, the backbone feature extraction network of YOLOv11n was reconstructed using a lightweight LNet-G design to minimize both the model's parameters and computational cost. Furthermore, we replaced YOLOv11n's original C3k2 neck module with a VoVGSCSP module, which reduces computation while improving detection accuracy. Ultimately, we substituted YOLOv11n's original Complete-IoU (CIoU) loss with the Weighted Intersection over Union (WIoU) variant, improving the localization precision of bounding boxes. Quantitative results obtained from the PV Evolution Labs (PVEL)-AD-2021 dataset indicate that LDV-YOLO achieves 1,783,282 parameters, 4.7 Giga Floating Point Operations Per second (GFLOPs), and 84.1% mAP. Compared to YOLOv11n, the proposed model reduces the parameter count by 31%, decreases GFLOPs by 25.40%, and improves mean average precision by 3.7%. The improved algorithm exhibits superior performance in PV cell defect detection while maintaining a lightweight architecture suitable for deployment on embedded devices.

Keywords: solar cells, lightweight, defect detection, YOLOv11, loss function

MSC: 68T20, 68T45

Abbreviation

CNN	Convolution Neural Network
CV	Computed Vision
AI	Artificial Intelligence
LN	Lightweight Network
FDD	Fast Defect Detection

1. Introduction

Monocrystalline Silicon Photovoltaic (PV) panels are critical components in renewable energy applications and have become the mainstream choice for solar cells due to their excellent performance and mature fabrication processes. However, during the manufacturing and operational phases of monocrystalline silicon PV panels, factors such as material defects, process variations, and environmental stresses often lead to various micro- and macro-scale defects, which adversely affect their photoelectric performance and long-term reliability. Manufacturing process defects can result in issues such as micro-cracks and gate line roughening, while harsh environmental conditions during prolonged outdoor operation may trigger failure modes like black core formation. Additionally, the material itself undergoes natural aging over extended service periods, typically manifested as finger degradation. These defect mechanisms cause substantial deterioration in the power generation efficiency of photovoltaic cells, thereby negatively impacting the overall capacity and reliability of power generation in solar power plants [1]. These flaws jeopardize operational safety while considerably degrading both the energy output and longevity of photovoltaic installations, thereby undermining the market competitiveness of renewable energy enterprises. Consequently, the development of high-performance defect inspection technologies for Mono-Si PV modules is of substantial practical importance in safeguarding plant operational security and enhancing the economic viability of the renewable energy sector.

In recent years, significant advancements in machine vision technology have emerged, particularly through the synergistic innovation of deep learning algorithms. These advancements have significantly enhanced both detection precision and processing speed in industrial quality inspection. Although electroluminescence imaging has become a standard practice for inspecting photovoltaic modules, the image analysis phase continues to rely heavily on manual interpretation. This reliance not only escalates inspection costs but also introduces considerable variability in detection outcomes. Consequently, the primary challenge in advancing this technology lies in achieving automated analysis to enhance both detection efficiency and result consistency.

With recent technological advancements, machine vision-based defect inspection systems are increasingly supplanting human-operated detection and traditional image analysis methods due to their remarkable improvements in throughput and favorable cost-benefit ratios. Contemporary object detection architectures are primarily categorized into two types: single-stage detectors, such as Solid State Drive (SSD) [2] and the You Only Look Once (YOLO) series, and two-stage detectors, including Region-Convolutional Neural Networks (R-CNN) and Faster R-CNN [3]. Although two-stage networks provide superior detection accuracy, they also present higher computational complexity, leading to slower inference speeds and increased memory consumption. For small target recognition applications, the SSD algorithm still has significant room for enhancement, with a noticeable performance gap when compared to the YOLO series. Experimental studies indicate that YOLO series models surpass mainstream object detection methods in terms of both recognition precision and processing latency. As the current iteration in the YOLO series, YOLOv11 not only retains the lightweight architectural advantages of its predecessors but also achieves a more favorable balance between detection accuracy and inference speed. Recent research has shown that the comprehensively optimized YOLOv11 demonstrates exceptional performance across various computer vision benchmarks, with enhancements primarily evident in: 1) the implementation of advanced network architectures; 2) the deployment of innovative feature integration schemes; and 3) the refinement of training methodologies. Collectively, these advancements have significantly enhanced YOLOv11's overall detection capabilities.

This study systematically analyzes and experimentally validates that the single-stage detection network YOLOv11 demonstrates substantial benefits for object detection tasks, primarily owing to its outstanding processing speed and minimal hardware demands. To address the unique challenges in photovoltaic module defect detection—such as diverse defect patterns, varying scales, and complex backgrounds—this study proposes an improved YOLOv11 architecture that is capable of more accurately identifying and detecting multiple imperfections in photovoltaic cells. The key innovations of this research include: 1) A redesigned feature extraction network is achieved through the Lightweight CPU Network-based GAM (LCNet-G), an optimized lightweight backbone for YOLOv11, thus decreasing model complexity and inference overhead. 2) The original C3k2 module in YOLOv11's Neck layer is substituted with a novel Vector of Vector Generalized Spatial pyramid Pooling Cross Stage Partial (VoVGSPCSP) architecture [4], which reduces computational overhead while improving

model accuracy. 3) The conventional Complete-IOU (CIOU) loss is substituted with Weighted Intersection over Union (WIoU) [5], improving convergence rate and localization precision.

2. Related work

In application scenarios such as real-time image processing and video analysis, the real-time performance and low-latency characteristics of models have become critical technical metrics. Particularly for edge devices, IoT endpoints, and embedded systems with limited resources, lightweight models exhibit superior computational efficiency, effectively meeting the stringent real-time processing requirements of these applications. The attention mechanism effectively focuses on key features while suppressing redundant information by assigning weights to input data. In scenarios involving the transmission of models or result data, this mechanism can significantly reduce

Machine vision-based inspection techniques have gained significant traction for solar cell surface defect detection in recent years. Su et al. [6] enhanced Faster R-CNN through synergistic channel-spatial attention mechanisms, enabling effective identification of three defect types in electroluminescence imagery. Su et al. [7] developed a Bidirectional Attention Fusion Pyramid Network (BAFPN-based) architecture for enhanced multi-scale feature integration. Su et al. [8] systematically evaluated EL defect detection performance comparing Faster RPN-CNN, BAF-Detector variants (D0-D2), and YOLOv5 architectures. Results showed that YOLOv5 delivered superior performance. Acikgoz et al. [9] developed an evolutionary deep learning framework for precise identification of seven defect categories in electroluminescence imagery. Wang et al. [10] successfully identified eight types of defect features in Experion Local (EL) images by introducing the Coordinate Attention (CA) mechanism into the feature extraction process and integrating a ResNet152-Xception hybrid architecture. Regarding algorithm optimization, Fu and Cheng [11] introduced the Experion Local Control Network (ELCN) module into the YOLOv7 framework in an innovative manner, significantly enhancing the detection accuracy for four categories of EL defects. Lu et al. [12] showed that combining the CA mechanism with the YOLOv5 algorithm greatly improved the model's performance, achieving outstanding results in detecting nine categories of defects. Furthermore, the research team of Li [13] made breakthrough progress in identifying four specific types of defects by integrating the Global Context Self-Attention (GCSc) mechanism with the YOLOv7 algorithm. Tella et al. [14] conducted a systematic assessment of multiple representative deep neural networks in the task of photovoltaic cell defect classification. In terms of model architecture innovation, Graph Interactive Transformer (GiT) [15] proposed a novel network architecture based on continuous interaction between graph structures and Transformers, achieving efficient collaborative representation of global and local features. Rodriguez et al. [16] introduced an innovative segmentation approach for tilted units and conducted a systematic evaluation of how different loss functions in the YOLOv5 network affect defect detection performance. For model optimization, Acikgoz [17] enhanced YOLOv7's backbone through ghost convolution and Global Attention Mechanism (GAM) integration. This design not only effectively improves the network's feature learning capacity but also substantially decreases redundant feature representations within the model. In addition, the Pamungkas research group [18] successfully developed a lightweight model for photovoltaic fault classification by integrating the U-Net architecture with the dense connection module of DenseNet. Aidan et al. [19] designed an improved YOLOv8 s-seg model based on the YOLOv8 framework. Tang et al. [20] developed a transfer learning approach for CNN depth reduction. This method innovatively integrates a multi-scale feature extraction module and was validated through experiments on the Visual Geometry Group (VGG)-16 architecture. The experimental findings demonstrate the approach's efficacy in model complexity reduction, but the overall accuracy of the model only increased by 0.98 percentage points. Li et al. [21] developed an infrared micro-defect detection method for photovoltaic modules, featuring a Multi-scale Parameter-free Edge Detection (MPED) module for enhanced small defect feature extraction, an improved SlimNet architecture for multi-scale feature fusion, and an Adaptive Auxiliary Detection Layer (AAD-Layer) enabling effective feature reuse. The proposed framework demonstrates competitive performance while maintaining computational efficiency. Chen et al. [22] presented a Gather-Distribute Domain Shift suppression network (GDDS) for open-world photovoltaic defect detection. The method employs a single-domain generalized architecture to overcome limitations of traditional dual-network approaches, integrating a DeepSpine module for multi-scale context

capture and background suppression. By combining cross-layer interactive learning for enhanced defect localization and normalized Wasserstein distance for detection optimization, GDDS demonstrates improved adaptability in open-environment defect detection scenarios. Yang et al. [23] designed a Lightweight Decoder-only DETR (LD-DETR) for photovoltaic module defect detection. By eliminating the Transformer encoder structure and employing a lightweight convolutional module to generate refined fusion features, the method significantly reduces computational costs while maintaining detection accuracy. Bao et al. [24] innovated CCA-YOLO with integrated channel-coordinate attention for PV defect detection. The method employs an the Residual Coordinate Convolution-based ECA (RCC-ECA) module to enhance feature representations, combines an Multiscale Defect Feature Localization Module (MDFLM) module for multi-scale detection robustness, and utilizes RCC-Up/Down sampling with the Residual Feature fusion with Coordinate Convolution-based CBAM (RFC-CBAM) fusion to maintain coordinate awareness. This approach significantly improves detection of irregular, multi-scale defects while maintaining computational efficiency for practical PV plant applications.

Existing research methods exhibit the following major technical limitations: First, convolutional operations with fixed receptive fields are significantly limited in their multi-scale feature extraction capacity under low computational constraints, making it difficult to effectively capture the morphological diversity of PV panel defects. Second, the fixed kernel structure used in conventional Spatial Pyramid Pooling-Fast (SPPF) modules lack sufficient sensitivity to boundary features, resulting in poor recognition of subtle defect edges. Moreover, although adding small-object detection layers or attention modules can improve the model's responsiveness to specific defect areas, this approach suffers from clear recognition bottlenecks when handling densely packed and adjacent small objects. It not only fails to accurately distinguish among neighboring small targets but also leads to increased false positives and missed detections. Additionally, it introduces extra computational overhead, significantly increasing post-processing time. Ultimately, these limitations hinder real-world implementation on resource-constrained platforms. These problems undermine the reliability and practicality of industrial applications, necessitating further research.

3. Methodology

3.1 LNet-G: design of a lightweight backbone network

PP-LCNet is a lightweight and high-performance convolutional neural network model designed by Cui et al. [25] in 2021. The architecture employs a cascaded arrangement of Depthwise Separable Convolution (DSC) building blocks, and the ReLU activation function within DSC is replaced with H-Swish to further boost performance. The DSC module integrates two key components: a Depthwise Convolution (DW) followed by Pointwise Convolution (PW). First, The DW operation independently processes each input channel with a $w \times w$ spatial filter, aggregating the results through channel-wise concatenation. However, The depthwise convolution preserves the original channel dimensionality. During the PW operation, 1×1 convolution kernels are used to change the number of output channels and perform channel fusion, ultimately generating a new feature map. For an input with three channels, the PW operation with a $w \times w$ convolution kernel and the DW operation with a 1×1 convolution kernel produces an output with M channels. The number of parameters Q in the depthwise separable convolution can be precisely represented by Formula (1), and its computation expression is as follows:

$$Q = 3 \times w^2 + 3 \times 1 \times 1 \times M \quad (1)$$

The parameter quantity K in conventional convolution layers is computable through Equation (2), expressed mathematically as:

$$K = w^2 \times 3 \times M \quad (2)$$

The ratio R of the number of parameters between depthwise separable convolution and standard convolution can be quantified using Formula (3), and its mathematical expression is defined as follows:

$$R = 1/M + 1/w^2 \quad (3)$$

Depthwise separable convolution achieves significant parameter reduction more effectively than standard convolution, and the use of pointwise convolution with larger convolution kernels further reduces the parameter scale. Therefore, this study constructs a network model based on depthwise separable convolution modules, achieving simultaneous optimization of both parameter count and computational cost.

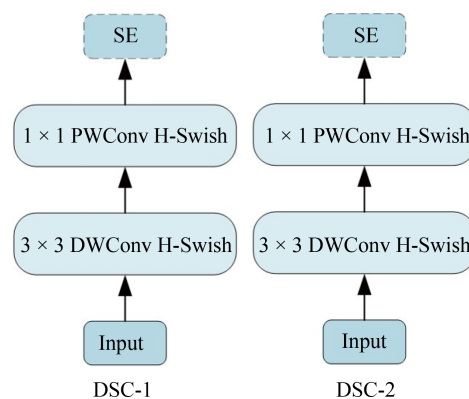


Figure 1. DSC network structure

Table 1. LCNet-G structure

Stage	Structure	Convolution kernel size	Stride	GAM
1	Conv2D	3×3	2	×
2	DepthSepConv	3×3	1	×
3	DepthSepConv	3×3	2	×
4	DepthSepConv	3×3	1	×
5	DepthSepConv	3×3	2	×
6	DepthSepConv	3×3	1	×
7	DepthSepConv	3×3	2	×
8	DepthSepConv	5×5	1	×
9	DepthSepConv	5×5	2	√
10	DepthSepConv	5×5	1	√

The PP-LCNet network incorporates two types of DSC module structures, as illustrated in Figure 1. The DSC-1 architecture sequentially combines 3×3 spatial filtering (DW) with 1×1 channel fusion (PW). The SE module, indicated by the dashed box, is an optional structure that improves the network's feature representation ability by utilizing a channel attention mechanism. The DSC-2 module extends the kernel size to 5×5 while maintaining the original architecture, with the remaining structural parameters consistent with those of the DSC-1 module. Addressing the SE module's restrictive channel-only attention mechanism, this study introduces the GAM module [26], which utilizes a dual-channel and spatial attention mechanism. This enhances global channel interaction, reduces information scattering, and amplifies global interaction representations to improve deep neural network capabilities. LCNet-G is designed based on this, significantly

improving the network's ability to extract features. The structure of LCNet-G is shown in Table 1. Stage 1 employs a 3×3 convolutional layer with stride-2 downsampling. Stages 2 to 10 consist of stacked DSC structures, aimed at reducing the network's parameter count and computational cost. In stages 8 through 10, 5×5 convolutions replace the 3×3 convolutions, significantly boosting the network's features.

Moreover, The GAM attention is incorporated within the final pair of DSC modules, which strengthens the global channel interaction and effectively improves the model's feature discrimination capability.

3.2 Lightweight based on VoVGGSCSP module

The introduction of a lightweight backbone inevitably compromises the model's accuracy in defect detection. To address this issue, GSConv [27], which represents a novel convolutional paradigm, has been adopted. GSConv integrates the lightweight characteristics of DSC with the accuracy of Standard Convolution (SC). This strategy not only facilitates effective reuse of feature information but also significantly reduces computational complexity. Thus, building on GSConv, Hu et al. [28] made enhancements by incorporating the GS bottleneck. Then, as illustrated in Figure 2, they adopted a one-shot aggregation strategy to construct the inter-stage subnetwork VoVGGSCSP module [29].

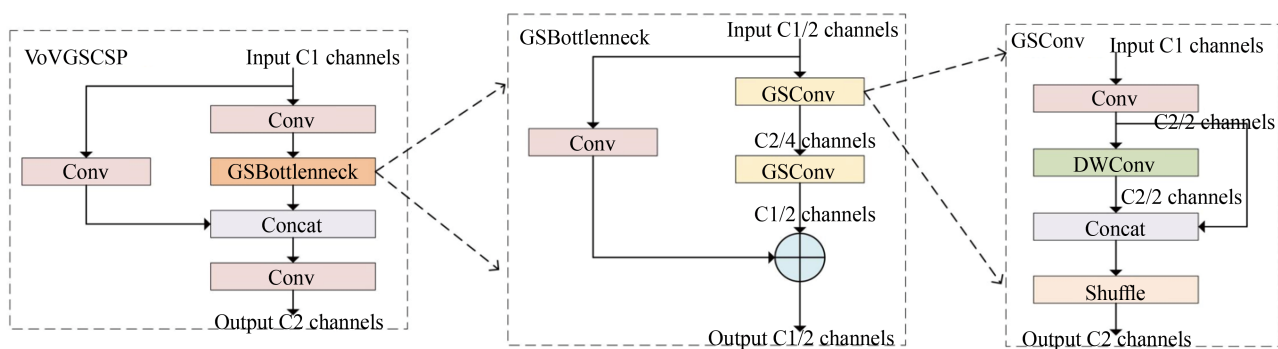


Figure 2. The structure of VoVGGSCSP

VoVGGSCSP improves upon GSConv through feature map bifurcation processing. This architecture implements a bifurcated processing flow: one processes features through a combination of Conv and GSConv structures, while the other applies a standalone Conv layer as a residual connection. Subsequently, The bifurcated features are then fused and projected to the output via a convolutional layer. The unique architecture of VoVGGSCSP enables easy dimensional operations and feature size reduction, thereby lowering computational load.

Within the deep learning framework, the C3k2 component in the neck is optimized using the VoVGGSCSP module. Model complexity is reduced alongside significant enhancement in micro-scale object identification accuracy. Importantly, GSConv is combined with the GSBottleneck to process feature maps characterized by maximum channel numbers and minimal spatial dimensions. At this stage, the feature maps exhibit minimal redundancy, eliminating the need for compression and allowing the module to operate more efficiently.

3.3 WIoU loss function

Addressing photovoltaic defect detection challenges characterized by cluttered backgrounds and prevalent micro-defects, we substitute the conventional CIoU loss with WIoU, enhancing small-target sensitivity and overall detection accuracy. The proposed method incorporates WIoUv3's dynamic Feature Modulation (FM) for enhanced loss computation. To achieve better generalization with minimal intervention during training, a distance attention mechanism is constructed, yielding WIoUv1's dual-layer attention architecture. Under this configuration, when the condition is met, the anchor box for regular samples is expanded, while for high-quality samples, it is reduced. Consequently, high anchor-target overlap prioritizes center-point alignment optimization. As shown in Equations (4) and (5):

$$L_{WIoU_{v1}} = R_{WIoU} L_{IoU} \quad (4)$$

$$R_{WIoU} = \exp \left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) \quad (5)$$

The inherent sample quality imbalance in photovoltaic cell datasets leads to inevitable inclusion of suboptimal training instances. As a result, WIoUv3 is developed on the foundation of WIoUv1, incorporating a dynamic non-monotonic FM, which effectively mitigating training-phase performance degradation factors. The proposed method demonstrates enhanced capability in micro-scale object identification. The WIoUv3 metric is formally defined as follows in Equations (6) and (7):

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \quad (6)$$

$$L_{WIoU_{v3}} = r L_{WIoU_{v1}}, \quad r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (7)$$

3.4 Network architecture

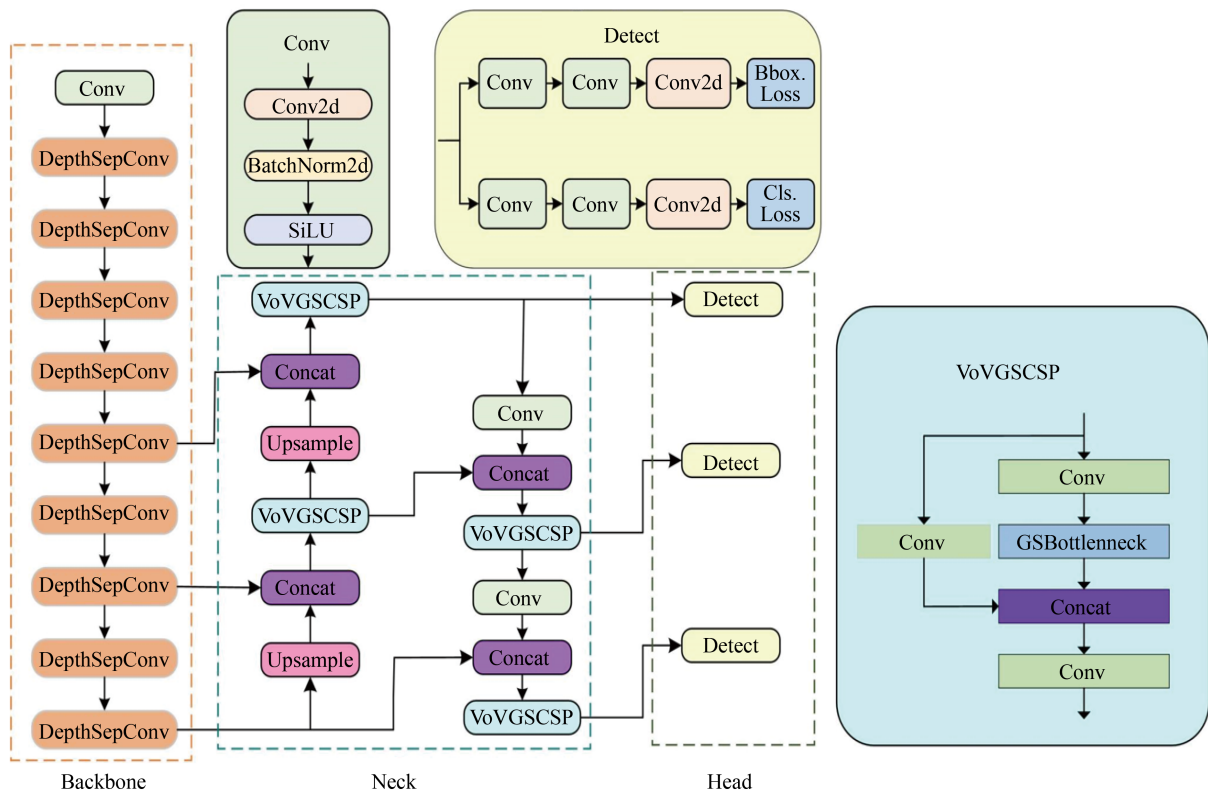


Figure 3. LDV-YOLO network architecture

An improved network structure is proposed to tackle the difficulties of small target defect detection and the large size of detection models in the context of Electroluminescence (EL) imaging for photovoltaic cells. The architecture described in Figure 3 has a significant effect in addressing this issue. Specifically, by designing the lightweight network LCNet-G to reconstruct YOLOv11n's core feature learning architecture, the network's parameters and computational load are reduced; the YOLOv11n architecture substitutes its Neck-layer C3k2 module with VoVGSCSP for enhanced feature fusion, which reduces computational load while improving the model's accuracy; finally, CIoU is ultimately superseded by WIoU in YOLOv11n, yielding superior localization accuracy and reduced omission errors. This improvement enables the model to remain lightweight while achieving good recognition accuracy, making it especially suitable for deployment tasks in photovoltaic defect detection.

4. Experiments and results

4.1 Dataset

This research employs two open-access photovoltaic defect datasets. The ELPV benchmark [30], developed by Buerhop-Lutz et al., comprises 2,624 eight-bit grayscale images (300×300 px) capturing diverse defect patterns across 44 PV modules with progressive performance degradation. Notably, the dataset lacks annotated defect labels. Instead, it classifies images into four categories based on defect probability: 0% (no defects), 33% (likely defects), 66% (likely defects), and 100% (defective) are shown in Figure 4.

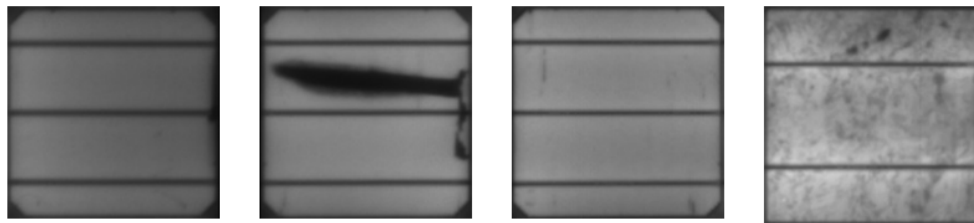


Figure 4. Defect confidence level = 0, defect confidence level = 1, defect confidence level = 0.33, defect confidence level = 0.66

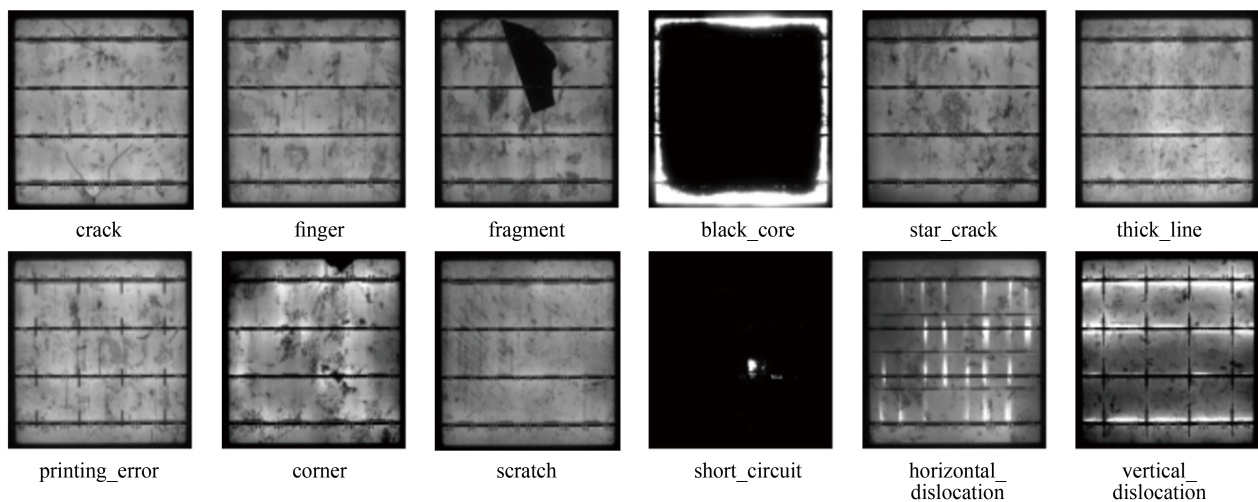


Figure 5. Representative samples demonstrating the 12 defect categories in the dataset

The PVEL-AD-2021 dataset was introduced by researchers from Hebei University of Technology and Beihang University [8]. The dataset encompasses 13 image categories, comprising normal references and 12 precisely annotated defect variants, with illustrative examples provided in Figure 5. An 8 : 1 : 1 split ratio was strictly maintained across training, validation, and test subsets during dataset partitioning to ensure rigorous experimental evaluation.

4.2 Experimental setting

All experiments were implemented on a Windows 10 workstation featuring an NVIDIA RTX A4,000 GPU (16 GB VRAM), Intel Xeon Gold 6248R CPU (3.0 GHz), and 768 GB RAM. The deep learning framework was built upon PyTorch 2.0.0 (CUDA 11.8) with Python 3.10.16. Training employed a batch size of 16 across 300 epochs. The model was optimized using SGD with cosine annealing, configured with: initial learning rate (η) = 0.01, momentum (μ) = 0.937, and L2 regularization (λ) = 0.0005.

4.3 Evaluation indicators

Deep learning object detection models are rigorously evaluated using standardized metrics that collectively assess detection accuracy, computational efficiency, and inference speed. The evaluation framework incorporates mean Average Precision (mAP) for comprehensive localization performance across varying IoU thresholds, with mAP@0.5 specifically measuring precision at the 50% overlap criterion. Model complexity is quantified through parameter count (in millions) and Floating-Point Operations (FLOPs), while practical deployment considerations are evaluated via per-image inference time measured in milliseconds.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P dR \quad (10)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (11)$$

In object detection evaluation, key metrics are defined as follows: for a dataset with N categories, precision (P) measures the proportion of true positives (TP , correctly detected objects) among all positive predictions ($TP + FP$, where FP denotes false positives), while recall (R) quantifies the fraction of true positives identified from all actual targets ($TP + FN$, with FN representing false negatives). Sample classification is determined by thresholding the Intersection-Over-Union (IoU) between predicted and ground-truth bounding boxes, where predictions exceeding the specified IoU threshold are considered positive samples.

4.4 Experiment results

4.4.1 Comparison of backbone improvement techniques

This study conducts comparative validation by substituting the YOLOv11n backbone in LDV-YOLO with alternative lightweight architectures to assess model efficacy. Table 2 shows the results of this experiment. The restructured LCNet-G backbone exhibits a marginal mAP reduction of 0.013%, compared to YOLOv11n, the parameter count decreased by 31% and the GFLOPs decreased by 25.40%, demonstrating significant advantages in terms of lightweight design.

Compared to MobileViTv3, HGNet, and PP-LCNet (SE), the mAP has increased, and the parameter size and GFLOPs are the smallest among them. Figure 6 provides a comparative visualization of mAP50 performance between LCNet-G and alternative backbone architectures, as well as the faster convergence and better stability of the model during training. Experimental results demonstrate the enhanced algorithm's dual capability: substantially compressing model dimensions while accelerating inference speed.

Table 2. Comparison results of backbone improvements

Model	Param (M)	FPS	GFLOPs	mAP0.5 (%)	mAP0.5 : 0.95 (%)
YOLOv11n (baseline)	2.58	68.5	6.3	80.9	51.8
MobileViTv3	2.29	50.9	5.4	76.8	49.7
HGNet	11.50	53.3	25.1	78.6	50.8
PP-LCNet (SE)	1.31	62.1	4.1	74.7	46.2
LCNet-G	1.27	56.6	3.0	79.6	50.3

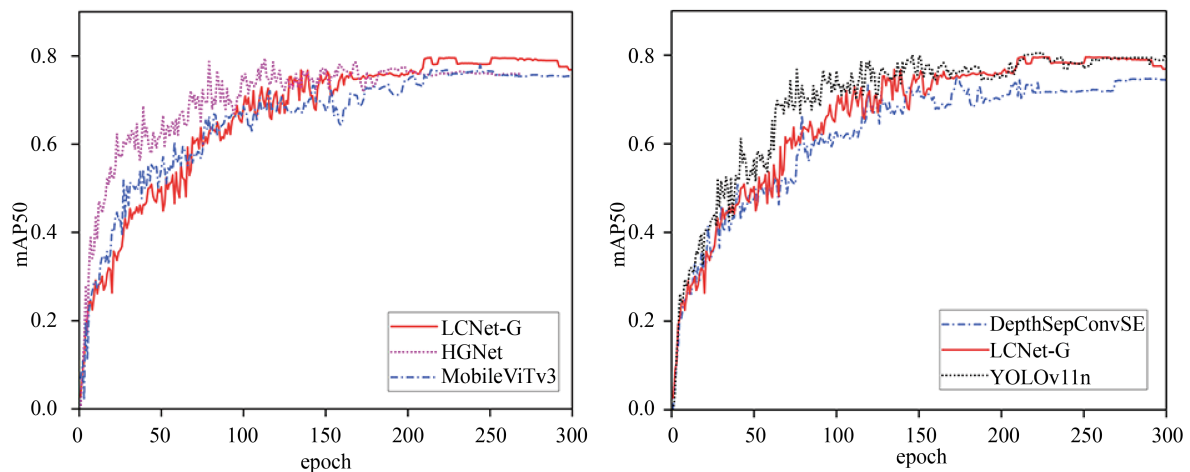


Figure 6. mAP50 results comparing different backbone improvements

4.4.2 Comparison experiment

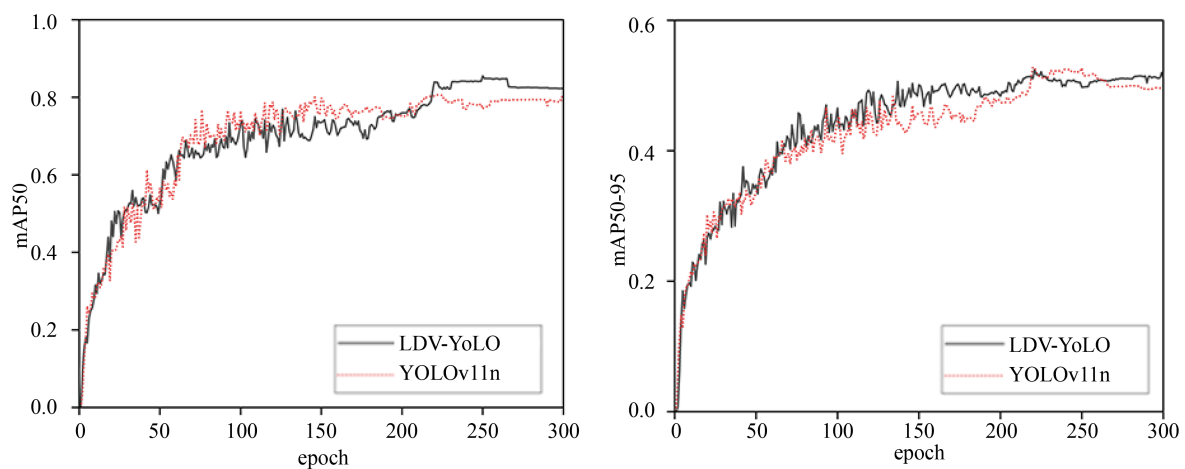
Building upon Su et al. [8] findings of YOLOv5's superior PV defect detection on PVEL-AD-2021, we conduct extended benchmarking against multiple architectures (including EfficientDet-D0, EfficientDet-D1, EfficientDet-D2, RT-DETR, YOLOv5n, YOLOv7-tiny, YOLOv8n, YOLOv9n, and YOLOv11n), with comprehensive results in Table 3.

During the training phase, all models were trained under identical hyperparameter configurations and dataset conditions, with input images resized to a fixed resolution of 640×640 pixels. The performance comparison curves were generated by applying appropriate smoothing to the metrics logged during training. As shown in Figure 7, both LDV-YOLO and YOLOv11 exhibit a rapid increase in mAP as training progresses. However, LDV-YOLO consistently achieves higher accuracy on the solar panel defect dataset, demonstrating the enhanced detection capability of the modified architecture.

To provide additional validation, we conducted supplementary training on the publicly accessible ELPV binary dataset, with the corresponding performance metrics detailed in Table 4.

Table 3. The statistical results of the ablation experiments

Model	mAP50	mAP50-95	Parameters	FPS	GFLOPs
EfficientDet-D0	0.491	0.231	3,874,217	10.2	5.23
EfficientDet-D1	0.564	0.321	6,578,346	14.5	9.43
EfficientDet-D2	0.632	0.396	8,194,523	20.1	13.6
RT-DETR	0.781	0.468	32,830,736	11.2	108.0
YOLOv5n	0.635	0.483	2,704,372	9.1	15.9
YOLOv7-tiny	0.541	0.385	6,044,754	13.3	105.3
YOLOv8n	0.806	0.548	2,686,708	68.3	6.8
YOLOv9n	0.804	0.531	2,751,848	70.4	39.7
YOLOv11n	0.809	0.518	2,584,492	68.5	6.3
LDV-YOLO	0.841	0.564	1,783,282	55.2	4.7

**Figure 7.** The detection performance metrics (mAP50 and mAP50-95) were systematically compared between the baseline YOLOv11 model and our novel detection framework**Table 4.** Detection results of our model on ELPV dataset

Model	mAP50	mAP50-95	Parameters	GFLOPs
YOLOv11	0.827	0.536	2,357,834	6.4
Ours	0.896	0.588	1,558,945	3.9

4.4.3 Ablation study

To systematically assess the performance improvements attributable to the LCNet-G module, VoVGSCSP module, and WIoU loss function integration in YOLOv11n, we performed comprehensive ablation studies. The detection results are shown in Table 5, where I represents the LCNet-G module, II represents the VoVGSCSP module, and III represents the WIoU loss function. Table 5 demonstrates that the YOLOv11n baseline model attains 80.9% mAP@0.5 and 51.8% mAP@0.5 : 0.95 for EL defect detection in PV cells, with a parameter count of 2.58 M and real-time processing capability of 68.5 FPS. After replacing the feature extraction network of YOLOv11n with LCNet-G, the mAP@0.5 and mAP@0.5 : 0.95 decreased by 0.13% and 0.15%, respectively, the FPS dropped by 11.1 frames/s, but the number of parameters was reduced by 1.31 M, and GFLOPs decreased by 3.3. The substitution of C3k2 with VoVGSCSP in YOLOv11n's

neck network improved $mAP@0.5$ by 1.2% and $mAP@0.5 : 0.95$ by 0.9%. However, FPS dropped by 23.3 frames/s, parameters increased by 0.12 M, and GFLOPs slightly decreased by 0.2. Replacing the CIoU loss function in YOLOv11n with WIoU led to a 0.9% and 0.2% increase in $mAP@0.5$ and $mAP@0.5 : 0.95$, respectively. Meanwhile, FPS decreased by 15.1 frames/s, the number of parameters dropped by 0.12 M, and GFLOPs were reduced by 0.8. Compared with YOLOv11n, LDV-YOLO achieved increases of 3.2% and 4.6% in $mAP@0.5$ and $mAP@0.5 : 0.95$, respectively, while FPS decreased by 13.3 frames/s, the number of parameters decreased by 0.8 M, and GFLOPs dropped by 1.6.

Table 5. The impact of individual architectural components on model detection performance

Model	Parameters	GFLOPs	FPS	$mAP@0.5$	$mAP@0.5-0.95$
YOLOv11n	2,584,492	6.3	68.5	0.809	0.518
YOLOv11n + I	1,270,410	3.0	57.4	0.796	0.503
YOLOv11n + II	2,611,204	6.1	45.2	0.821	0.527
YOLOv11n + III	2,461,580	5.5	53.4	0.818	0.520
YOLOv11n + I + II	1,683,282	4.2	43.1	0.834	0.533
YOLOv11n + I + III	1,370,410	3.5	58.4	0.812	0.474
YOLOv11n + II + III	2,448,676	5.7	44.8	0.833	0.547
YOLOv11n + I + II + III	1,783,282	4.7	55.2	0.841	0.564

The ablation study results demonstrate that our enhanced model maintains detection accuracy while achieving model lightweighting, providing a practical solution for PVEL defect detection hardware implementation.

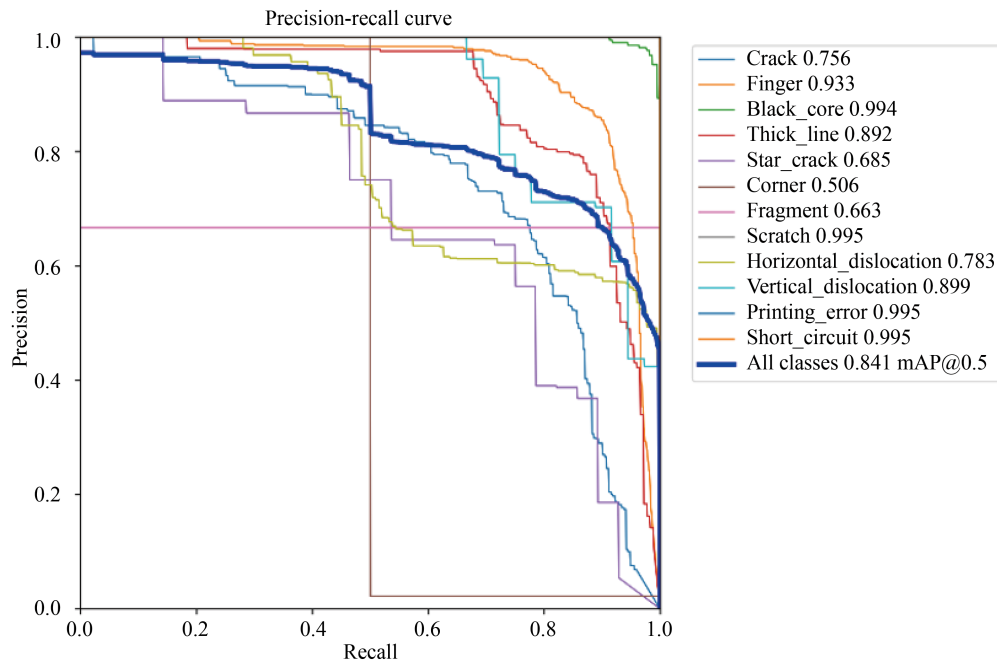


Figure 8. Precision-recall curve for the LDV-YOLO

The model's robust detection capability for multiple common defect types (finger, black core, scratch, printing error, short circuit, and vertical dislocation) is demonstrated in Figure 8, with industrial-grade $mAP50$ performance being achieved. However, for defects like cracks, thick lines, and horizontal dislocations—due to their visual similarity—the

model exhibits only moderate mAP50 performance, indicating room for improvement. Defects like star cracks, corner breakages, and fragments show mAP50 values under 70%, suggesting that additional enhancement of detection accuracy is still required. The observed performance degradation primarily stems from high background interference, resulting in elevated false-positive and false-negative rates, especially for small-scale defects. Incorporating supplementary features or adopting alternative algorithms may help improve the model's ability to extract distinguishing characteristics for these challenging defects.

Lastly, The height correspondence between predicted and ground truth bounding boxes is visually confirmed in Figure 9, showing precise vertical alignment post-training.

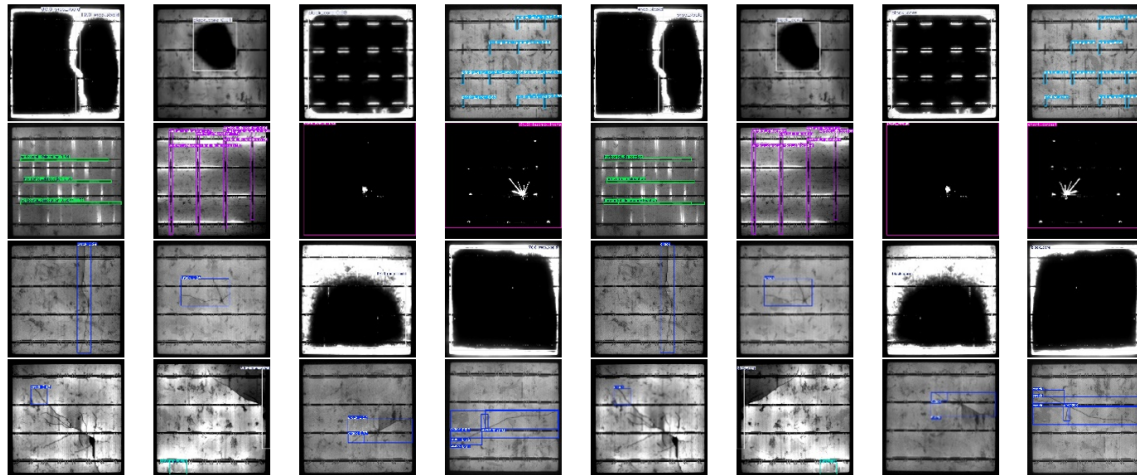


Figure 9. A comparative analysis was performed between predicted bounding boxes and randomly selected ground truth annotations

5. Conclusions and future directions

This study proposes a lightweight LDV-YOLO model to address the challenges of solar cell surface defect detection, including difficulties in industrial deployment, high computational requirements, and limited detection accuracy. Through the following innovations, the model achieves lightweight design while improving detection performance.

1. **Lightweight backbone:** The substitution of the conventional YOLOv11n backbone with LCNet-G achieves a 50.7% reduction in model parameters and 52.3% decrease in computational complexity without compromising detection accuracy, enabling practical implementation in embedded systems for photovoltaic defect inspection.

2. **Optimized VoVGSCSP architecture:** In order to handle multi-scale defects and enhance detection accuracy without sacrificing inference speed, architectural modifications involved substituting the neck layer's C3k2 module with our VoVGSCSP implementation.

3. **Improved loss function:** To accelerate convergence during training, the CIoU loss function in the original model was replaced with WIoU.

According to both comparison and ablation experiments, the proposed LDV-YOLO model reduced parameters by 31%, GFLOPs by 25.40%, and achieved a 3.7% increase in mAP over the original model. These results demonstrate that the model is well-suited for practical industrial applications.

5.1 Future directions

Future work will focus on three key improvements:

1. Data Augmentation: Addressing class imbalance through advanced data augmentation techniques like DCGAN and M-mask augmentation, particularly for underperforming defect categories (corner, fragment, and star_crack) which currently show sub-70% mAP despite an overall 84.1% mAP.

2. Enhanced network architecture: Enhancing small defect detection by optimizing the network's multi-scale feature extraction capabilities, potentially through refined feature pyramid structures or attention mechanisms; and

3. Lightweight optimization: Implementing model compression techniques including knowledge distillation, network pruning, and quantization to achieve industrial-grade real-time performance while maintaining detection accuracy, thereby bridging the gap between laboratory results and practical deployment requirements. This comprehensive optimization strategy aims to simultaneously improve detection performance for challenging defect types while meeting stringent computational efficiency demands for photovoltaic quality inspection applications.

By addressing these challenges and extending the method's capabilities, this approach has the potential to significantly enhance defect detection accuracy, further reduce model parameters, and ultimately achieve industrial-grade applicability.

Conflict of interest

The authors declare no conflicts of interest.

References

- [1] Herraiz AH, Marugán AP, Márquez FPG. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. *Renewable Energy*. 2020; 153: 334-348. Available from: <https://doi.org/10.1016/j.renene.2020.01.148>.
- [2] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, et al. SSD: Single shot multibox detector. In: *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands; October 11–14 2016; Proceedings, Part I 14*. Cham: Springer International Publishing; 2016. p.21-37.
- [3] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016; 39(6): 1137-1149. Available from: <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [4] Li R, He Y, Li Y, Qin W, Abbas A, Ji R, et al. Identification of cotton pest and disease based on CFNet-VoV-GCSP-LSKNet-YOLOv8s: a new era of precision agriculture. *Frontiers in Plant Science*. 2024; 15: 1348402. Available from: <https://doi.org/10.3389/fpls.2024.1348402>.
- [5] Tong Z, Chen Y, Xu Z, Yu R. Wise-IOU: bounding box regression loss with dynamic focusing mechanism. *Computer Science*. 2023. Available from: <https://doi.org/10.48550/arXiv.2301.10051>.
- [6] Su B, Chen H, Chen P, Bian GB, Liu W. Deep learning-based solar-cell manufacturing defect detection with complementary attention network. *IEEE Transactions on Industrial Informatics*. 2020; 17(6): 4084-4095. Available from: <https://doi.org/10.1109/TII.2020.3008021>.
- [7] Su B, Chen H, Zhou Z. BAF-detector: An efficient CNN-based detector for photovoltaic cell defect detection. *IEEE Transactions on Industrial Electronics*. 2021; 69(3): 3161-3171. Available from: <https://doi.org/10.1109/TIE.2021.3070507>.
- [8] Su B, Zhou Z, Chen H. PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection. *IEEE Transactions on Industrial Informatics*. 2022; 19(1): 404-413. Available from: <https://doi.org/10.1109/TII.2022.3162846>.
- [9] Acikgoz H, Korkmaz D, Budak U. Photovoltaic cell defect classification based on integration of residual-inception network and spatial pyramid pooling in electroluminescence images. *Expert Systems with Applications*. 2023; 229: 120546. Available from: <https://doi.org/10.1016/j.eswa.2023.120546>.
- [10] Wang J, Bi L, Sun P, Jiao XG, Ma X, Lei XY, et al. Deep-learning-based automatic detection of photovoltaic cell defects in electroluminescence images. *Sensors*. 2022; 23(1): 297. Available from: <https://doi.org/10.3390/s23010297>.

- [11] Fu H, Cheng G. Convolutional neural network based efficient detector for multicrystalline photovoltaic cells defect detection. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*. 2023; 45(3): 8686-8702. Available from: <https://doi.org/10.1080/15567036.2023.2230935>.
- [12] Lu S, Wu K, Chen J. Solar cell surface defect detection based on optimized YOLOv5. *IEEE Access*. 2023; 11: 71026-71036. Available from: <https://doi.org/10.1109/ACCESS.2023.3294344>.
- [13] Li J, Wu W, Chen H. GCSC-detector: A detector for photovoltaic cell defect based on deep learning. In: *2023 42nd Chinese Control Conference (CCC)*. Piscataway, NJ: IEEE; 2023. p.6913-6917.
- [14] Tella H, Mohandes M, Liu B, Rehman S, Al-Shaikhi A. Deep learning system for defect classification of solar panel cells. In: *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*. Piscataway, NJ: IEEE; 2022. p.448-453.
- [15] Shen F, Xie Y, Zhu JQ, Zhu XB, Zeng HQ. Git: Graph interactive transformer for vehicle re-identification. *IEEE Transactions on Image Processing*. 2023; 32: 1039-1051. Available from: <https://doi.org/10.1109/TIP.2023.3238642>.
- [16] Rodriguez AR, Holicza B, Nagy AM, Vörösházi Z, Bereczky G, Czúni L. Segmentation and error detection of PV modules. In: *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)*. Piscataway, NJ: IEEE; 2022. p.1-4.
- [17] Acikgoz H. An automatic detection model for cracks in photovoltaic cells based on electroluminescence imaging using improved YOLOv7. *Signal, Image and Video Processing*. 2024; 18(1): 625-635. Available from: <https://doi.org/10.1007/s11760-023-02724-7>.
- [18] Pamungkas RF, Utama IBKY, Jang YM. A novel approach for efficient solar panel fault classification using coupled udensenet. *Sensors*. 2023; 23(10): 4918. Available from: <https://doi.org/10.3390/s23104918>.
- [19] Barrett A, Bratanov D, Amarasingam N, Sera D, Gonzalez F. Machine learning based damage detection in photovoltaic arrays using UAV-acquired infrared and visual imagery. In: *2024 International Conference on Unmanned Aircraft Systems (ICUAS)*. Piscataway, NJ: IEEE; 2024. p.264-271.
- [20] Tang R, Ren Z, Ning S, Dezso S, Felipe G. Fault classification of photovoltaic module infrared images based on transfer learning and interpretable convolutional neural network. *Solar Energy*. 2024; 276: 112703. Available from: <https://doi.org/10.1016/j.solener.2024.112703>.
- [21] Li W, Li J, Cao B, Zhu J, Tian MH. FAA-YOLO: A method for defects detection of small infrared targets in photovoltaic modules. *IEEE Sensors Journal*. 2025; 25(6): 10486-10497. Available from: <https://doi.org/10.1109/JSEN.2025.3534277>.
- [22] Chen H, Zhang Y, Zhang Y, Yan XW, Zhang X, Zou KL. Defect detection of photovoltaic panels to suppress endogenous shift phenomenon. *IEEE Transactions on Semiconductor Manufacturing*. 2024; 38(1): 83-95. Available from: <https://doi.org/10.1109/TSM.2024.3510358>.
- [23] Yang Y, Zhang J, Shu X, Pan L, Zhang M. A lightweight transformer model for defect detection in electroluminescence images of photovoltaic cells. *IEEE Access*. 2024; 12: 194922-194931. Available from: <https://doi.org/10.1109/ACCESS.2024.3520239>.
- [24] Bao J, Yuan X, Wu Q, Lam CT, Ke W, Li P. CCA-YOLO: Channel and Coordinate Aware-Based YOLO for photovoltaic cell defect detection in electroluminescence images. *IEEE Transactions on Instrumentation and Measurement*. 2025; 74: 1-12. Available from: <https://doi.org/10.1109/TIM.2025.3541805>.
- [25] Cui C, Gao T, Wei S, Du YN, Guo R, Dong S, et al. PP-LCNet: A lightweight CPU convolutional neural network. *Computer Science*. 2021. Available from: <https://doi.org/10.48550/arXiv.2109.15099>.
- [26] Shen S, Yang J. Better YOLO with attention-augmented network and enhanced generalization performance for safety helmet detection. *Computer Science*. 2024. Available from: <https://doi.org/10.48550/arXiv.2405.02591>.
- [27] Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *Computer Science*. 2022. Available from: <https://doi.org/10.48550/arXiv.2206.02424>.
- [28] Hu S, Lou Z, Yan X, Ye Y. A survey on information bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2024; 46(8): 5325-5344. Available from: <https://doi.org/10.1109/TPAMI.2024.3366349>.
- [29] Wang Z, Liu F. Research on remote sensing target detection algorithm based on improved YOLOv5. In: *Fourth International Conference on Machine Learning and Computer Application (ICMLCA 2023)*. Bellingham, WA: SPIE; 2024. p.135-140.

- [30] Grisanti M, Spatafora MAN, Ortis A, Battiato S. E-ELPV: extended ELPV dataset for accurate solar cells defect classification. In: *Proceedings of SAI Intelligent Systems Conference*. Cham: Springer Nature Switzerland; 2023. p.837-848.